DOCUMENT RESUME

ED 205 551 TM 810 430

AUTHOR Yamomoto, Yoneo; Wise, Steven L.
TITLE Extracting Undimensional Chains from Multidimensional Datasets: A Graph Theory Approach.
INSTITUTION Illinois Univ., Urbana. Computer-Based Education Research Lab.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO CERL-RR-80-2
PUB DATE Feb 80
CONTRACT N00014-79-C-0752
NOTE 34p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Algorithms; Data Processing; Mathematical Applications; Testing
IDENTIFIERS Data Sets; *Graph Theory; Internal Consistency; Matrix Operations; *Order Analysis

ABSTRACT

An order-analysis procedure, which uses graph theory to extract efficiently nonredundant, unidimensional chains of items from multidimensional data sets and chain consistency as a criterion for chain membership is outlined in this paper. The procedure is intended as an alternative to the Reynolds (1976) procedure which is described as being exhaustive in the number of computer calculations it requires for chain extraction. Order analysis concepts are discussed and graphically represented before the general consistency index used in the procedure is introduced and defined. Procedures for chain extraction follow, in which it is stated that the dominance matrix can be reconsidered as a labeled digraph. From this digraph, all subgroups with perfect consistency are generated, and these in turn are used as starting points in the chain extraction process. The original dominance digraph is then reduced until the chain is found from each subgraph. The graph-theoretic algorithm may be carried out using a series of matrix manipulations performed on the dominance matrix. An illustrative example of the method is provided. In conclusion, possible extensions of the method are suggested. (Author/AEF)

Computer-based Education Research Laboratory

University of Illinois Urbana Illinois

EXTRACTING UNIDIMENSIONAL CHAINS FROM MULTIDIMENSIONAL DATASETS: A GRAPH THEORY APPROACH

YONEO YAMAMOTO

STEVEN L. WISE

This research was sponsored by the Personnel and Training Research Program, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-79-C-0752. Contract Authority Identification Number NR 150-415.

COMPUTERIZED ADAPTIVE TESTING AND MEASUREMENT RESEARCH REPORT 80-2

FEBRUARY 1980

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report No. 80-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Extracting Unidimensional Chains from Multidimensional datasets: A Graph Theory Approach		5. TYPE OF REPORT & PERIOD COVERED Aug. 21 - Nov 20, 1979 6. PERFORMING ORG. REPORT NUMBER CERL Report E-14
7. AUTHOR(s) Yoneo Yamamoto & Steven L. Wise		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0752
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer-based Education Research Labor tory University of Illinois Urbana, IL 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR042-04; RR042-04-01; NR154-445
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE February 1980 13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

order analysis, dimensionality, graph theory, internal consistency, testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Reynolds (1976) described an order-analysis procedure for extracting unidimensional chains of items (or persons) from multidimensional datasets. It is an exhaustive method, using one of Cliff's consistency indices to extract all possible chains which maintain a given level of consistency. A practical problem with this method is that the number of computer calculations needed for chain extraction can become very high.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

This paper outlines a procedure which uses graph theory to extract nonredundant chains efficiently, rather than exhaustively generating all chains, as with Reynold's procedure. It also uses chain consistency as a criterion for chain membership. The dominance matrix can be reconsidered as a labelled digraph. From this digraph, all subgroups with perfect consistency are generated, and these in turn are used as starting points in the chain extraction process. The original dominance digraph is then reduced until the chain is found from each subgraph. This graph-theoretic algorithm may be carried out using a series of matrix manipulations performed on the dominance matrix.

ACKNOWLEDGEMENTS

We wish to acknowledge the services rendered by the following people:

Kumi and Maurice Tatsuoka, for their helpful editorial comments and suggestions.

Robert Baillie, whose expert computer programming was invaluable in the operationalization of our algorithm.

Wayne Wilson for his talented artwork.

Louise Brodie for her clerical assistance.

ABSTRACT

Reynolds (1976) described an order-analysis procedure for extracting unidimensional chains of items (or persons) from multidimensional datasets. It is an exhaustive method, using one of Cliff's consistency indices to extract all possible chains which maintain a given level of consistency. A practical problem with this method is that the number of computer calculations needed for chain extraction can become very high.

This paper outlines a procedure which uses graph theory to extract nonredundant chains efficiently, rather than exhaustively generating all chains, as with Reynold's procedure. It also uses chain consistency as a criterion for chain membership. The dominance matrix can be reconsidered as a labelled digraph. From this digraph, all subgroups with perfect consistency are generated, and these in turn are used as starting points in the chain extraction process. The original dominance digraph is then reduced until the chian is found from each subgraph. This graph-theoretic algorithm may be carried out using a series of matrix manipulations performed on the dominance matrix.

EXTRACTING UNIDIMENSIONAL CHAINS FROM MULTIDIMENSIONAL DATASETS: A GRAPH THEORY APPROACH

Yoneo Yamamoto & Steven Wise[1]

INTRODUCTION

Many psychometricians have investigated the problem of extraction of factors or dimensions from a data matrix. Ordering theory (Airasian and Bart, 1972) is one such method. Order analysis (Krus, Bart, & Airasian, 1975) was developed to examine the logic-based dimensions in binary data matrices. It makes use of the dominance relations (Coombs, 1964) in data to determine simple orders.

Cliff (1977) developed several order consistency indices from the dominance relation. He shows his indices to be comparable to classical measures of internal consistency such as the Kuder-Richardson formula and Loevinger's index of homogeneity.

Reynolds (1976) developed an order analysis procedure which uses one of Cliff's consistency indices to recover all possible chains (i.e., all possible dimensions that maintain a given level of consistency). A practical problem with this method is that the number of computer calculations needed for chain extraction can easily become very high.

This article is concerned with presenting an efficient algorithm to extract all possible chains using a general consistency index. Our extraction procedure is an interactive method using the PLATO computer system at the University of Illinois. The main algorithm uses graph theory to extract nonredundant chains efficiently, rather than

---

[1]The authors wish to acknowledge the contribution of Robert Baillie, who wrote computer programs for the routines developed in this paper.

exhaustively generating all chains as with Reynolds' (1976) procedure.

ORDER ANALYSIS

The methods described in this paper are based on the concepts of order analysis (Krus, Bart, & Alrasian, 1975; Krus, 1975). One advantage of using order analysis is that it is more easily restated in terms of graph theory concepts than are other multivariate procedures such as factor analysis and cluster analysis. In the following section, we will explain the notation to be used in describing the extraction algorithm.

Simple Orders

Mathematically, simple orders are defined as follows. Let R denote a simple order relation between elements of a set A. The following three properties hold for all elements a, b, and c of set A:

1) Asymmetry property

   aRb implies $b\bar{R}a$, where $\bar{R}$ means "not R".

2) Transitivity property

   aRb and bRc implies aRc.

3) Connectedness property

   either aRb or bRa.

An example of simple order is shown by the concept "less than". Let A be the set {1, 2, 3, 4}. Then the set R of all ordered pairs of elements from A which maintain the relation "less than" is given by $R_1 = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$.

Directed Graphs

Graph theory is a useful tool for helping one to better understand complicated sets of relationships (particularly logical

relationships). We will show that graphs can also be useful in developing algorithms. We will use directed graphs to extract elements from sets which maintain the properties of simple orders.

Let us denote a directed graph G by G = (V,E), where V is the set of vertices and E the set of edges (ordered pairs of vertices). For example, if we let V_1 be the set A described earlier,

$G_1 = (V_1, E_1)$

$V_1 = A = \{1,2,3,4\}$

$E_1 = R_1 = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$

G_1 can be represented graphically, as shown in Figure 1.

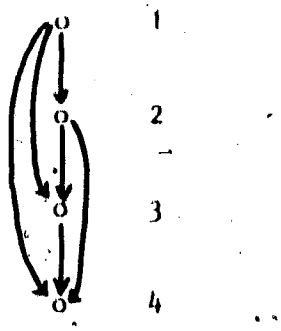

Figure 1. Graph for G_1

Notice that the three properties of simple orders (asymmetry, transitivity, and connectedness) are shown graphically.

Matrix Representation of Directed Graphs

In order to represent a directed graph G in a computer, one commonly uses a V × V adjacency matrix (Deo, 1974) where g_{ij} = 1 or 0 depending on whether or not a directed edge exists between vertices i and j in V. The matrix representation for G_1 is shown in Figure 2.

$$Q_1 = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix}$$

Figure 2. Matrix Representation of Q_1

In testing, one is typically dealing with a persons-by-items data matrix. Figure 3 shows a (5 x 4) data matrix, S.

$$\begin{array}{c} \\ \text{Persons} \end{array} \begin{array}{c} \begin{matrix} \text{Items} \\ a \quad b \quad c \quad d \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} = S$$

Figure 3. A Persons by Items Data Matrix

The element s_{ij} of S is equal to 1 if person i gets item j correct, and 0 otherwise. In order analysis terms, it is said that when $s_{ij} = 1$, person i dominates item j (Krus, 1975). Conversely, when $s_{ij} = 0$, item j is said to dominate person i. The data matrix S is representative of a perfect Guttman scale. When the items are ordered in terms of difficulty, each person will get each item correct up to a given difficulty level,

and no items correct thereafter. In actual testing, however, perfect Guttman scales are rarely found.

Dominance Matrices

In order analysis, one attempts to determine sets of items (or equivalently, persons) that exhibit the properties of simple orders. To find these orders, one constructs a dominance matrix from the raw data matrix. Following Cliff (1975) a supermatrix A, containing S and S' (the transposed complement of S) as its segments, is defined. The elements of S and S' contain information regarding the dominances between the sets ofpersons and items. The lack of initial dominance information about elements within each set is denoted by zero matrices in the supermatrix A, which is known as an adjacency matrix.

$$A = \begin{bmatrix} 0 & S' \\ S & 0 \end{bmatrix}$$

However, we are ultimately concerned precisely with the dominance relations between elements within sets. If we multiply A by itself, we obtain the matrix A^2, which contains the within-set dominance information.

$$A^2 = \begin{bmatrix} N & 0 \\ 0 & X \end{bmatrix}$$

The submatrix N is the item dominance matrix. Element n_{ij} of N is equal to the number of persons for which item i dominated item j (that is, item i was incorrect and item j was correct). Correspondingly, the submatrix X of A^2 is the person dominance matrix. Element x_{st} equals the number of items for which person s dominated person t (that is, the number of items which person s got correct and t got incorrect).

At this point, the order analysis would concentrate on either R or R and extract chains (subsets) of items or persons which exhibit the properties of simple orders to a sufficient degree of approximation. The number of chains needed to account for all of the elements in the set is interpreted as the dimensionality of the set (Airasian, 1975).

GRAPHIC REPRESENTATION OF DOMINANCE INFORMATION

This section outlines the graphic representation of several order-analysis concepts described in the previous section. Labelled directed graphs will be used in a later section to develop an extraction algorithm.

Item-Person Dominance Relations

Recall the binary, items-by-persons data matrix S from the last section. In Figure 4a the arrow from i to j denotes that person i dominates item j ($s_{ij} = 1$). Figure 4b shows that item j dominates person i ($s_{ij} = 0$).

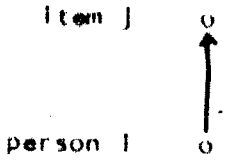

Figure 4a
Person i Dominates Item j

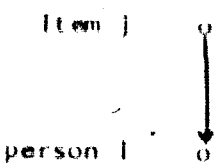

Figure 4b
Item j Dominates Person i

Item-Item Dominance Relations

Consider a person X responding to two items, a and b. Since person X can get each item either right or wrong, there are four possible

response patterns.

Item	
a	b
1	1
1	0
0	1
0	0

Only the (1,0) and (0,1) response patterns contain order information, as shown by the graphs in Figure 5. Paths of length 2 are found only for these two response patterns. (A path of length 2 means a succession of two arrows, one starting at the other's endpoint.)

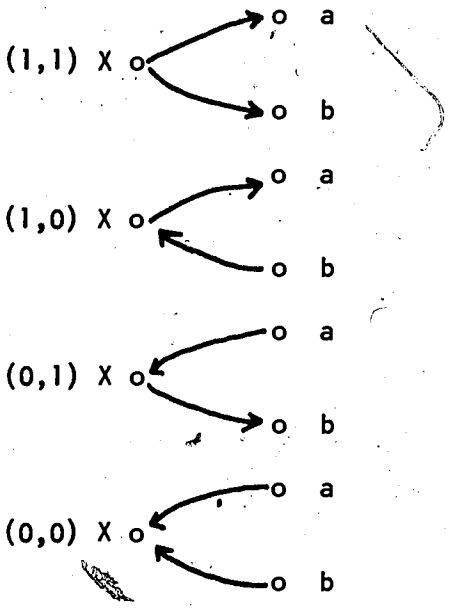

Figure 5. Possible Two Item Response Patterns and their Graphs

For an entire persons-by-items data matrix S, the item-person dominances are easily represented using a bipartite graph. A bipartite graph depicts the relations between the elements of one set and the

elements of another set. Figure 6 shows a data matrix, its adjacency matrix, and its corresponding bipartite graph.

$$S = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$A = \begin{bmatrix} 0 & \overline{S}' \\ S & 0 \end{bmatrix} = \begin{array}{c} \\ A \\ B \\ C \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} A & B & C & 1 & 2 & 3 & 4 \end{array} \\ \left[\begin{array}{ccc:cccc} 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hdashline 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

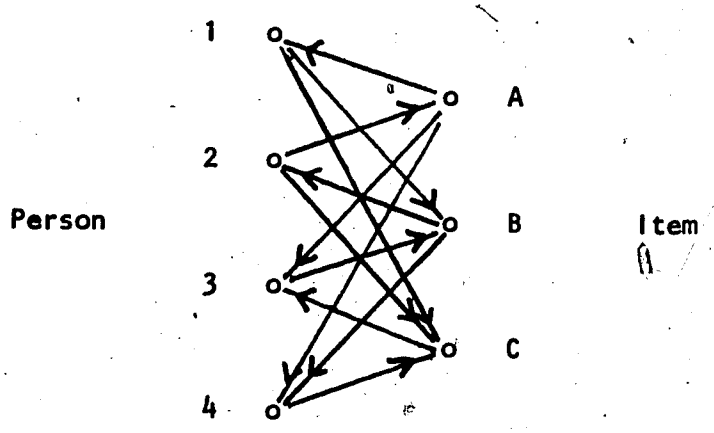

Figure 6. A Data Matrix with its Adjacency Matrix and Bipartite Graph

The item dominance matrix for the data matrix in Figure 6 can also be shown by use of a labelled graph. Figures 7a and 7b show this dominance matrix and its graph. An element n_{ij} represents the number of ij dominances. An element n_{ji} represents the number of ji dominances, and can also be thought of as the number of ij counterdominances. In the graph, the numerals in parentheses are related to these dominances and counterdominances as follows: The second number is the number of dominances in the direction of the referenced arrow, and the first number is the sum of the dominances and counterdominances.

$$N = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 2 & 2 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

Figure 7a. Item Dominance Matrix for Data Matrix in Figure 6.

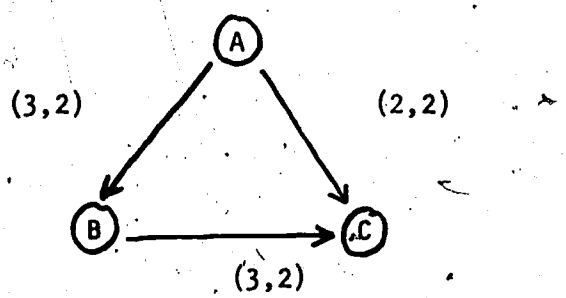

Figure 7b. Labelled Graph for Item Dominance Matrix

CONSISTENCY INDEX

We define a general consistency index, C, as

$$C = \alpha(U_a/U) - \beta$$

where $U = \sum_j \sum_k n_{jk}$ (the total number of item dominances)

$U_a = \sum_j \sum_{k>j} n_{jk}$ (the number of dominances above the main diagonal)

α and β are parameters which determine the scale and origin, respectively, of C.

For a given data matrix, the total number of dominances (U) is fixed, while U_a is dependent on the order of the items. When the items are ordered in terms of difficulty, U_a is maximized.

When $\alpha = 2$ and $\beta = 1$, C is equivalent to the index used by Reynolds in his exhaustive method. This is also equivalent to C_{t1}, one of Cliff's consistency indices (Cliff, 1977), since

$$C = \frac{2U_a}{U} - 1 = \frac{2U_a - U}{U} = \frac{U_a - (U - U_a)}{U} = \frac{U_a - U_b}{U} = \frac{v_m}{v} = C_{t1}$$

where $U_b = \sum_j \sum_{k>j} n_{jk}$ (the number of dominances below the main diagonal)

$v_m = \sum_j \sum_{k>j} (n_{jk} - n_{kj})$

$v = \sum_j \sum_k n_{jk}$

It is also easy to show that with $\alpha = 4$ and $\beta = 3$, C becomes C_{t2}, another index proposed by Cliff.

Althought the above equation for C involves two parameters, α and β, it is seen that β is a function of α by considering a dataset which forms a Guttman simplex so that $U_a = U$ and $C = 1$. Thus, $1 = \alpha(1) - \beta$, or $\beta = \alpha - 1$. Hence

$$C = \alpha(U_a/U) - (\alpha - 1)$$

Therefore, we have only one parameter α, and its value is determined so that C will be approximately equal to zero for a dominance matrix based on a random set of data. With appropriate choices of α, C can approximate Cliff's C_{t3}, C_{t4}, or C_{t5}.

Let $C' = \frac{U_a}{U} = \frac{C + \beta}{\alpha}$. Then :

$$C' = \frac{U_a}{U} = \frac{U_a}{U_a + U_b} = \frac{\sum_j \sum_{k>j} n_{jk}}{\sum_j \sum_{k>j} n_{jk} + \sum_j \sum_{k<j} n_{ik}} = \frac{\sum_j \sum_{k>j} n_{jk}}{\sum_j \sum_{k<j} (n_{kj} + n_{jk})}$$

$$= \frac{\sum_j \sum_{k>j} d_{jk}}{\sum_j \sum_{k<j} (C_{jk} + d_{jk})}$$

where d_{jk} is the number of dominances from j to k.

c_{jk} is the number of counterdominances from j to k $(=n_{kj})$.

Or we can say

$$C' = \frac{\sum_j \sum_{k>j} d_{jk}}{\sum_j \sum_{k<j} d_{jk}} = \frac{\sum_j \sum_{k>j} d_{jk}}{\sum_j \sum_{k>j} a_{kj}}$$

where $a_{jk} = c_{jk} + d_{jk}$.

Krus (1975) showed that dominance matrices can be split into additive matrices according to item or person subsets. From this, letting each person be treated as a subset, we have

$$C' = \frac{\sum_{j}\sum_{k>j}\sum_{i=1}^{n} \bar{s}'_{jk}\, s_{ik}}{\sum_{j}\sum_{k<j}\sum_{i=1}^{n} (\bar{s}'_{ki}\, s_{ij} + \bar{s}'_{ji}\, s_{ik})} = \frac{\sum_{j}\sum_{k>j}\sum_{i=1}^{N} d_{jk}^{(i)}}{\sum_{j}\sum_{k<j}\sum_{i=1}^{N} a_{jk}^{(i)}}$$

$$= \frac{\sum_{i=1}^{N}\sum_{j}\sum_{k>j}^{n} d_{jk}^{(i)}}{\sum_{i=1}^{N}\sum_{j}\sum_{k<j}^{n} a_{jk}^{(i)}} = \frac{\sum_{i=1}^{N} d^{(i)}}{\sum_{i=1}^{N} a^{(i)}}$$

This general consistency index, C, has the advantage of additivity over the above-mentioned indices C_{t3}, C_{t4}, and C_{t5}. Each item or person can be thought of as having an individual C value. When groups of items or persons are combined, both the numerator and denominator of the U_a/U term of C change by an additive amount. This is not the case for C_{t3}, C_{t4}, or C_{t5}, which are further influenced by "chance" consistency in a non-additive fashion.

In the next section we utilize this additive property to extract all possible unidimensional chains in a more efficient fashion.

EXTRACTION PROCEDURE

Our extraction procedure is an interactive method using the PLATO computer system at the University of Illinois. The main algorithm uses graph theory to extract nonredundant chains efficiently, rather than exhaustively generating all chains, as with Reynolds' (1976) procedure. It uses the consistency index C as a criterion for chain membership, taking into account the fact that the numerator and denominator both maintain additivity under addition and deletion of elements in a chain.

The dominance matrix can be interpreted as a labelled digraph. From this graph, all subgroups with perfect consistency are generated, and these in turn are used as starting points in chain extraction. Given a criterion consistency value C_o, the original dominance digraph is then reduced until a chain is found from each subgraph. After extracting all possible chains, redundant chains can be eliminated and C_o can be changed to select more or fewer chains, as desired.

As a backdrop, we first briefly review Reynolds' (1976) method, which extracts all possible chains which satisfy the condition that criterion C (which is equivalent to Cliff's consistency index C_{t1}) exceed some cutoff value. His approach is an iterative method generating all chains starting at each item.

For each item, k, a consistency value C_{p+k} is calculated for combining the new item with items already in the chain. The item which produces the smallest decrement in C_{p+k} is added to the chain, producing C'_p :

$$C'_p = \max_k C_{p+k} ,$$

This process of adding the "most consistent" item is repeated using the remaining items until the overall chain consistency drops below the criterion consistency value. At this point the procedure stops and the chain is complete. This method is reapplied using each item as a starting point so that, for m items, m chains are extracted. Duplicate chains are deleted, leaving a unique set of chains that are interpreted as the dimensions of the data matrix (Reynolds, 1976).

Reynolds' procedure tolerates a lot of redundancy in order to ensure that all chains are extracted. That is, for each item in a given chain, using it as a starting point will frequently result in the re-production of an already existing chain.

The total number of possible consistency calculations for a k-item data matrix can be computed. Starting at each item there are

$$(k-1) + (k-2) + \ldots + (k-k-1) = \frac{k(k-1)}{2}$$

calculations. Summed across starting points (items), we have a total of

$$k \times \frac{k(k-1)}{2} = \frac{k^2(k-1)}{2}$$

calculations, which is of the order of k^3. As k increases the number of calculations needed goes up rapidly. For instance, for a 20-item data matrix, as many as 3800 calculations may be needed. Many of these calculations are unnecessary.

* * *

Our algorithm, on the other hand, begins by determining all subsets of items with perfect consistency (C=1.0); that is, subsets

which satisfy

$$C_p = \frac{\sum_{p\varepsilon I} d_i}{\sum_{p\varepsilon I} a_i} = 1.0$$

where p denotes any chain of items with C = 1. (A chain consisting of only one item is permissible.) Note that there will likely be more than one starting item chain. At this point, items are added one by one to a given chain in such a way that its consistency stays as high as possible. When its value drops to some predetermined C_o, we stop adding further items to this chain, but select another chain and repeat the procedure.

We have considered three different strategies for deciding on the successive items to add to a chain:

(1) Take each one of the remaining items in turn, calculate the new consistency index when each is added to the chain, and select that item which yields the largest C value. This is essentially the same as Reynolds' (1976) procedure, and it is time consuming because repetitious calculations have to be done for each item in turn.

(2) Without taking the chain as a whole into consideration, look at the "individual" consistency index

$$C_{ind} = \frac{d_i}{a_i} \qquad (i \notin I)$$

of each remaining item, and choose the item with the largest C_{ind} to add to the chain. (Note, however, that each time an item is added to a chain, the individual consistency index for each of the remaining items has to be calculated anew.)

(3) Out of the remaining items, take the three with the highest individual consistency values, and add each one of these, in turn, to the chain. Choose the one that produces the smallest decrease in the overall chain consistency to actually add to the chain. (This method is to allow for the possibility -- which we have not experienced so far -- that the item with the highest individual consistency, selected by method (2), may not necessarily produce the smallest decrease in the chain consistency.)

Our computer algorithm can use any of the three strategies just described; moreover, even when method (1) is used, we have made arrangements to keep the repetitious calculations at a minimum. We shall now describe our algorithm in detail, giving an example using the second strategy.

Chain Extraction Algorithm

1. Start with a persons × items data matrix.

1.1 Compute the dominance matrix.

$$N = \left[n_{ij} \right] \qquad k,j = 1,2,\ldots,m$$

Note that this matrix is not a binary matrix and that $n_{ii} = 0$. Reorder the dominance matrix in terms of item difficulties.

1.2 Construct a matrix O such that each element in its lower triangle is equal to the sum of the corresponding element of N plus the symmetric element of N. The upper triangle elements are equal to the corresponding element of N and the main diagonal elements are all zero. That is,

$$O = [o_{ij}] \qquad i,j = 1,2,\ldots,m$$

$$\text{where } o_{ij} = \begin{cases} n_{ij} & i < j \\ 0 & i = j \\ n_{ij} + n_{ji} & i > j \end{cases}$$

2. Graph Initialization

2.1 Construct a labelled graph from matrix O.

$G = (V,E)$

$V = \{1,2,\ldots,m\}$ (the total set of items)

$E = \{(i,j) \quad i=1,2,\ldots,m;\ j=i+1,i+2,\ldots,m\}$

2.2 Initialize labels for vertices and edges:

$e_{ij} = (a_{ji}, d_{ij}) = (o_{ji}, o_{ij})$ for $i=1,2,\ldots,m;\ j=i+1,i+2,\ldots,m$

$v_i = [p_i, q_i] = [0,0]$ for $i=1,2,\ldots,m$

The graph labels refer to the consistencies both between vertices and "within" vertices. As the algorithm proceeds, some of the vertices will be merged to form new vertices. Each v_i will refer to the consistency of the set of items within vertex i. Hence, each p_i is equal to the sum of the elements in the lower triangle of O that correspond to the dominances between elements contained in vertex i. Likewise, q_i is equal to the sum of the corresponding elements of the upper triangle of O. Thus, each v_i is initialized at $[0,0]$.

2.3 Initialize a chain matrix (CH)

This will start out as an $m \times m$ binary matrix that will contain the item chains. All elements of CH will be initialized at zero except for the main diagonal elements, which are equal to one. When

the algorithm is finished, ch_{ij} will equal one if item j is a member of chain I; otherwise, it will be equal to zero. Also, redundant chains will be eliminated during the extraction process, so that the final number of chains (rows) in CH will be less than m.

3. Extraction of maximal chain with a consistency greater than a criterion, C_o'.

3.1 For a given chain L, merge the vertices and edges included in the chain. For vertices i and j (i<j) where $ch_{Li}=1$ and $ch_{Lj}=0$, merge vertices i and j to form vertex i•j with edges

$$\left.\begin{array}{ll}(k, i\cdot j) & k < i \\ (i\cdot j, k) & i < k\end{array}\right\} \quad \text{for all } k \in V-\{i,j\}.$$

The new labels in the graph are:

$$v_{i\cdot j} = [p_i + p_j + a_{ji},\ q_i + q_j + d_{ij}]$$

$$\left.\begin{array}{ll} e_{k,i\cdot j} = (a_{ik} + a_{jk},\ d_{ki} + d_{kj}) & k < i \\ e_{i\cdot j,k} = (a_{ki} + a_{jk},\ d_{ik} + d_{kj}) & i < k < j \\ e_{i\cdot j,k} = (a_{ki} + a_{kj},\ d_{ik} + d_{jk}) & k > j \end{array}\right\} \text{for all } k \in V-\{i,j\}.$$

3.2 Look for the candidate vertex (item) which should be merged next. Find the item with the largest "individual" consistency (strategy 2)

$$C_k' = \begin{cases} \dfrac{[e_{k,i\cdot j}]_2}{[e_{k,i\cdot j}]_1} & k < i \\ \dfrac{[e_{i\cdot j,k}]_2}{[e_{i\cdot j,k}]_1} & k > i \end{cases}$$

where $[e_{a,b}]_1$ and $[e_{a,b}]_2$ are the first and second components of $e_{a,b}$, and i•j represents items already in the chain.

3.3 Compute the new consistency (C') with the best candidate item, k, added to the chain.

$$C'_{k,i\cdot j} = \frac{q_{i\cdot j} + d_{i\cdot j,k}}{p_{i\cdot j} + a_{k,i\cdot j}} \qquad k < i$$

or

$$C'_{i\cdot j,k} = \frac{q_{i\cdot j} + d_{k,i\cdot j}}{p_{i\cdot j} + a_{i\cdot j,k}} \qquad k > i$$

3.4 If $C' \geq C'_o$ then add vertex k to the chain and merge the graph. Return to step 3.2 .

If $C' < C'_o$, the algorithm is finished.

AN ILLUSTRATIVE EXAMPLE

We will now illustrate the use of our algorithm using the example given in Reynolds (1976, p. 24). The consistency index he used is equivalent to our general consistency index with parameters $\alpha = 2$ and $\beta = 1$.

1. Data Matrix

1.1 The ordered item dominance matrix for the six items is:

$$N = \bar{S}' \times S =$$

	F	D	E	C	A	B
F	0	3	2	4	4	5
D	1	[illegible]		1	4	3
E	0	2	0	3	3	4
C	1	0	2	0	4	2
A	0	2	1	3	0	3
B	0	0	1	0	2	0

1.2 The corresponding 0 matrix is:

$$0 =$$

	F	D	E	C	A	B
F	0	3	2	4	4	5
D	4	0	2	1	4	3
E	2	4	0	3	3	4
C	5	1	5	0	4	2
A	4	6	4	7	0	3
B	5	3	5	2	5	0

2. Graph Initialization

2.1, 2.2 The labelled graph corresponding to matrix 0 is shown below, in Figure 8.

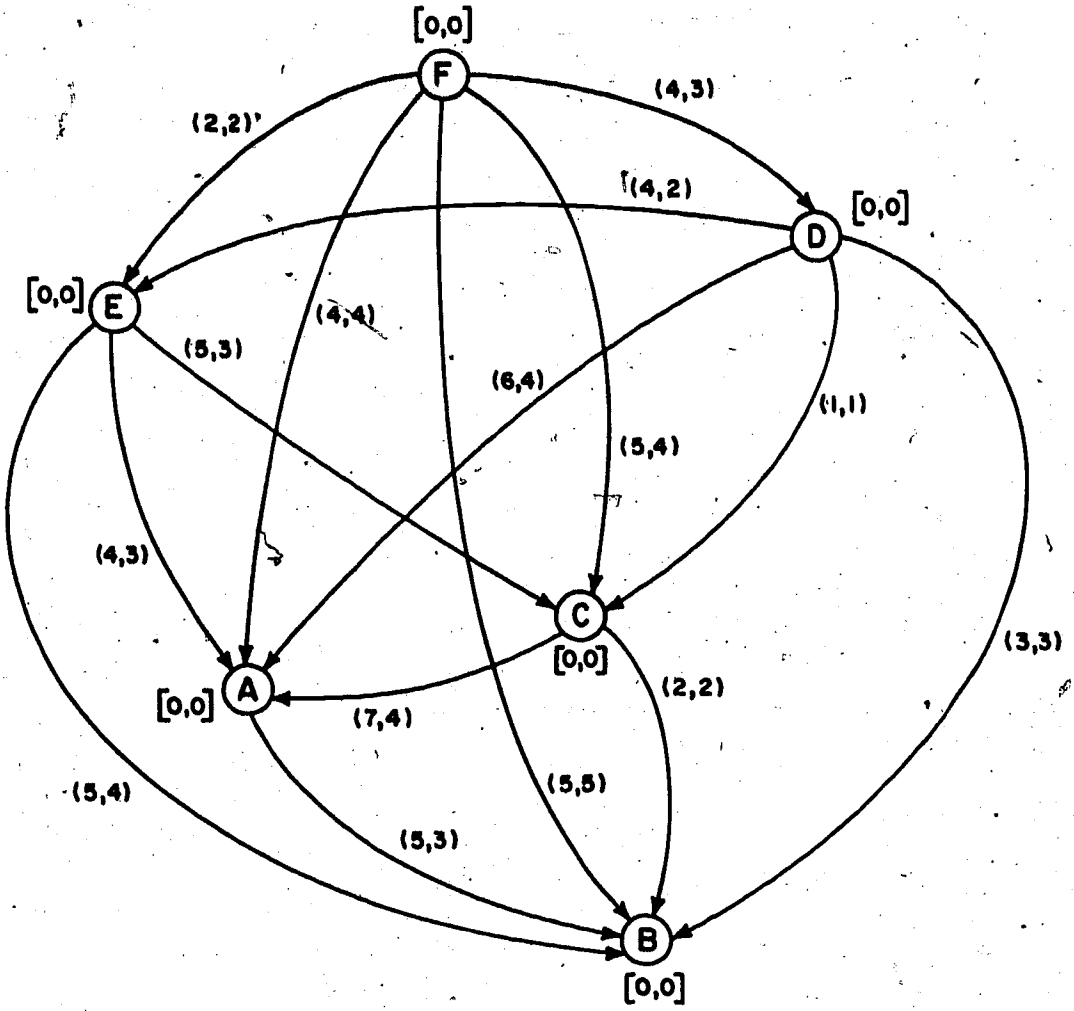

Figure 8. Labelled graph of matrix 0.

2.3 Initialize the chain matrix (CH).

$$CH = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} F & D & E & C & A & B \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

3. Extraction of Perfect Item Chains (C = 1.0) .

3.1, 3.2 Select vertices with perfect consistency (C = 1.0) for each chain, and delete redundant chains.

$$CH = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} F & D & E & C & A & B \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \quad \text{reduce} \quad CH = \begin{array}{c} \\ 1 \\ 2 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} F & D & E & C & A & B \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Repeat step 3. for a lower criterion consistency.

4. Extraction of Maximal Chain greater than a Criterion Consistency (C_o=.8).

4.1 Merge vertices and edges included in the chain.

We get Figure 9. Note that chain 1 is being processed.

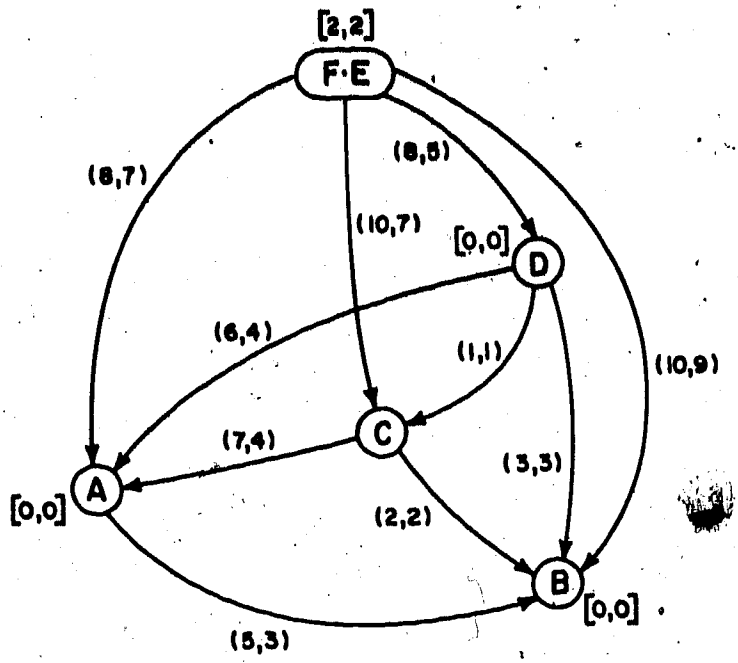

Figure 9. Merging of Vertices for Chain 1

4.2 Determine the next candidate item.

Of these 5/8, 7/10, 7/8, and 9/10, the value 9/10 is the largest. Therefore the vertex B is the candidate.

4.3 Compute C ($\alpha = 2$, $\beta = 1$) for the chain with the candidate added.

$C = 2(2+9)/(2+10)-1 = .8333$

4.4 If $C \geq C_o$ then add candidate item to chain, merge the graph and return to step 3.2. If $C < C_o$, do not add candidate item and chain is complete. $C\ (=.8333) \geq C_o\ (=.8)$ so we add the candidate, merge the graph and get CH matrix and Figure 10.

$$CH = \begin{array}{c} \\ 1 \\ 2 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} F & D & E & C & A & B \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

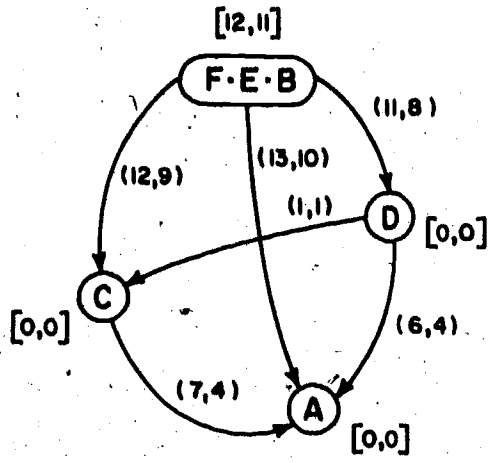

Figure 10. Merging of Vertex F, E and B

We repeat the 4.2 through 4.4. This time vertex A becomes the candidate because 10/13 is the largest value, but the new $C = 2(11+10)/(12+13)-1 = .68 < C(=.8)$, so the algorithm then stops.

Repeating the process for the other chains, we get the final chain matrix ($C_o = .8$).

$$CH = \begin{array}{c} \\ 1 \\ 2 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} F & D & E & C & A & B \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{l} .5161 \\ .8333 \\ .8000 \\ .8000 \end{array}$$

The value in the upper right-hand corner (.5161) is the C value for a chain containing all of the items. Note that each individual chain represents an improvement over .5161.

The interactive procedure to extract chains is summarized in Figure 11. Each block consists of one or more routines which are implemented on the PLATO system at the University of Illinois.

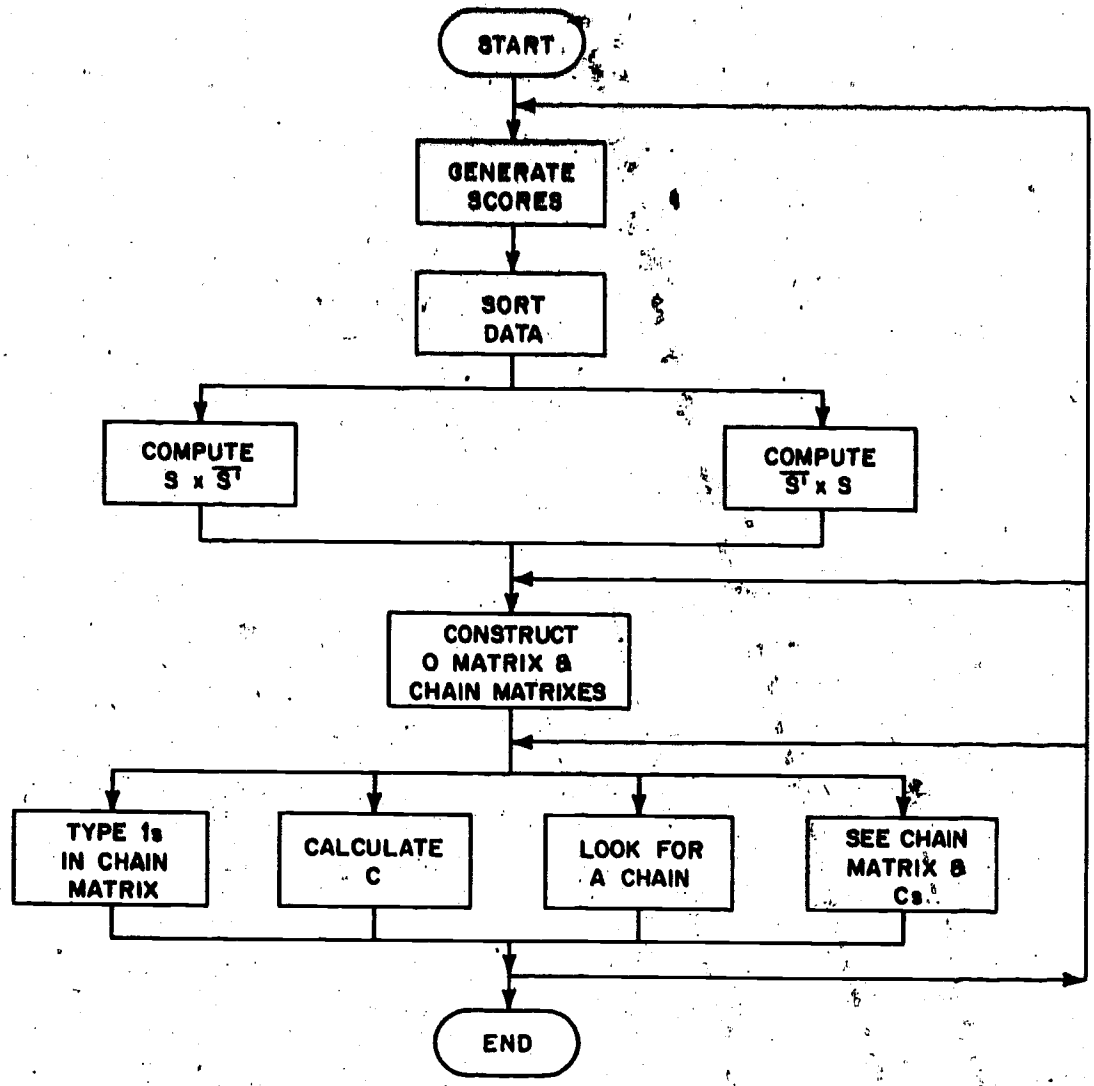

Figure 11. Extracting procedure on PLATO system.

DISCUSSION

We introduced a general consistency index in this paper such that, with parameters values $\alpha = 2$ and $\beta = 1$, it is equivalent to the index which Reynolds (1976) used in his article, and also to one of Cliff's indices, C_{t1} (1977). With $\alpha = 4$ and $\beta = 3$ it becomes C_{t2}, another index by Cliff. With appropriate choice of α (and $\beta = \alpha-1$) we can approximate others of Cliff's indices, C_{t3}, C_{t4}, and C_{t5}.

One important feature of the general consistency index is that additivity holds in both numerator and denominator. It enables us to use a graphical merging technique. As a result we can give an efficient algorithm to get all possible chains, which were obtained by Reynolds by his exhaustive method.

In the current paper we have picked up the items in order to get item chains. But we can easily extend the technique described here to a method that both picks up items and eliminates persons in order to get highly consistent item chains, because additivity in numerator and denominator holds for persons as well as items. In some cases a few persons "contaminate" our data, therefore elimination of these persons is a good way to get good item chains. Our efficient algorithm ensures the extracting of chains for a large number of items and persons. We can also apply the same technique to get consistent person chains, by picking up or eliminating items or persons.

REFERENCES

Airasian, P. & Bart, W. Ordering theory: A new and useful measurement model, Educational Technology, 1972, 13, 56-60.

Cliff, N. Complete orders from incomplete data: Interactive ordering and tailored testing, Psychological Bulletin, 1975, 82(2), 289-302.

Cliff, N. A theory of consistency of ordering generalizable to tailored testing, Psychometrika, 1977, 42(3), 375-399.

Coombs, C. H. A theory of data. New York: Wiley, 1964.

Coombs, C. H., Dawes, R. M. & Tversky, A. Mathematical psychology: An elementary introduction, Englewood Cliffs, N. J.: Prentice-Hall, 1970.

Deo, M. Graph theory with applications to engineering and computer science, Prentice-Hall, 1974.

Krus, D. J., Bart, W. M. & Airasian, P. W. Ordering theory and methods, Los Angeles: Theta Press, 1975.

Reynolds, T. J. The analysis of dominance matrices: Extraction of unidimensional orders within a multidimensional context, (NR No. 150-373, Tech. Rep. No. 3), Los Angeles: University of Southern California, 1976.